Machine Learning
Programming Assignment

Exercise 1 : File read and write

Irirs database - The data set contains 3 classes of 50 instances
each, where each class refers to a type of iris plant. It contains 4 numeric features and
the predicted attribute is class of iris plant.

Observation: None of the values were 0, hence didn't skip any index

Exercise 2 : Regression

The relation between x and y is non-linear. Hence to catch non-linear
relationship we generate Feature matrix=[1, x, x^2 , . . . , x^10]

Loss function = $\|\mathbf{Xw} - \mathbf{Y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2$ (L$_2$ norn)

Regularisation parameter = λ

- Idea is to first find best λ based on dividing input data to validation and training dataset.
- Next with best λ find the parameters
- Last check the accuracy on the training set by given parameter

Validation data = $1/4^{th}$ of the data for validation
Training data = $3/4^{th}$ of the data for training

Tried different lambda from $1e^{-4}$ to 10 on same validation set.
Best lambda is coming out to be 2.8676, validation_loss = 144.051

As we increase lambda our training loss gets higher (2.77 -> 10.86), but our
validation_loss decreases sharply (394.424 -> 144.051) because we are making our
model more rigrous to new data.

Loss on full training dataset is coming out to be 8.182